

028  
414  
s. 3360-91  
992







HD28  
.M414  
no.

3360-  
91-  
1992

The Investment Performance of U.S. Equity Pension  
Fund Managers: An Empirical Investigation

T. Daniel Coggin     \*  
Frank J. Fabozzi     \*\*  
Shafiqur Rahman     \*\*\*

WP #3360-91 EFA

Revised  
November 1992



The Investment Performance of U.S. Equity Pension  
Fund Managers: An Empirical Investigation

T. Daniel Coggin     \*  
Frank J. Fabozzi     \*\*  
Shafiqur Rahman     \*\*\*

WP #3360-91 EFA

Revised  
November 1992

\*     Virginia Retirement System  
\*\*    Journal of Portfolio Management  
\*\*\*   Portland State University

To be presented at the Annual Meeting of the American Finance Association, Anaheim, CA, January 1993. We thank Jon Christopher of the Frank Russell Company for providing the pension manager data used in this study. The paper has benefitted from discussions with John E. Hunter and Charles Trzcinka. The opinions and conclusions offered in this study do not necessarily represent those of the Virginia Retirement System or the Frank Russell Company.

MIT LIBRARIES  
DEC 07 1992  
RECEIVED



# **The Investment Performance of U.S. Equity Pension Fund Managers:**

## **An Empirical Investigation**

### **ABSTRACT**

This paper presents an empirical investigation of the security selection and market timing performance of a random sample of 71 U.S. equity pension fund managers using monthly returns for the period 1983-1990. The 71 equity fund managers include banks, insurance companies and investment advisors who have been allocated funds by pension plan sponsors. The data were provided by the Frank Russell Company of Tacoma, WA. While there have been many studies of U.S. equity mutual funds, ours is the first such study of which we are aware of U.S. equity pension fund managers. The estimates of selectivity and timing were derived using the Treynor and Mazuy (1966) model and the Bhattacharya and Pfleiderer (1983) model. The total sample of managers is subdivided into four groups by investment style, and a benchmark portfolio is identified for each style. We also included two benchmarks for the broad equity market. Regardless of the choice of a benchmark portfolio or estimation model, the selectivity measure is positive on average and the timing measure is negative on average. However, both selectivity and timing do appear to be somewhat more sensitive to the choice of a benchmark portfolio (and, possibly, the time period) when managers are classified by investment style. A meta-analysis was performed to quantify the effect of sampling error on the cumulated regression results. In every case, meta-analysis revealed some real variation (in excess of that attributable to sampling error) around the mean values for both selectivity and timing. An examination of the 80% probability intervals for selectivity revealed that the best managers can deliver substantial risk-adjusted excess returns. Finally, consistent with previous studies of equity mutual fund performance, we also found a negative correlation between selectivity and timing. However, we argue that the observed negative correlation in our data is largely an artifact of negatively correlated sampling errors for the two estimates.



# **The Investment Performance of U.S. Equity Pension Fund Managers:**

## **An Empirical Investigation**

Each year Pensions & Investments, a leading trade newspaper for the pension management industry, profiles the top 1000 public and private U.S. pension funds. At year-end 1990, these funds had total pension assets of \$1.876 trillion. Approximately \$750 billion (40 percent) was invested in equities. The Investment Company Institute estimates that \$250 billion was invested in open- and closed-end equity-oriented U.S. mutual funds at year-end 1990. This snapshot indicates a 3:1 ratio for pension fund equity investment versus mutual fund equity investment. Not only is the dollar difference large, but also the difference in the number of managers in each universe is large. The total number of pension fund managers is much larger than the number of mutual fund managers, by a ratio of approximately 10:1. Yet surprisingly little research has been done on the investment performance of U.S. equity pension fund managers. This paper begins to fill an important gap in the literature by providing empirical evidence on the investment performance of these managers.

The focus of this study is on equity pension fund managers who have been allocated funds by a pension plan sponsor. Brinson, Hood and Beebower (1986), Ippolito and Turner (1987), and Berkowitz, Finney and Logue (1988) examined the investment performance of a sample of large U.S. pension plans. Each plan may be composed of many fund managers in different asset categories with their own specific investment objectives and styles. In a recent study containing a wealth of information about the pension management industry, Lakonishok, Shleifer and Vishny (1992) examined the annual returns of a sample of equity pension funds over

the period 1983-1989. However, they made no risk adjustment, used only the S&P 500 Index as a benchmark portfolio, and focused on actual (equity only) returns (before management fees). Hence their results are not comparable to ours. To date, ours is the only study we know of which specifically examines the components of the investment performance of a sample of U.S. equity pension fund managers.

The two components we examine are security selection skill and market timing skill. Security selection involves the identification of individual securities which are under- or overvalued relative to the market in general. Within the specification of the Capital Asset Pricing Model (CAPM), the investment manager attempts to identify securities with expected returns which lie significantly off the security market line. The manager will then invest in those securities which offer an abnormally high risk premium. Market timing refers to forecasts of return on the market portfolio. If the manager believes he can forecast the market return, he will adjust his portfolio risk level accordingly.

According to the efficient market hypothesis, all active investment management activity is futile. The only rational investment choice for a plan sponsor is to invest in a passively managed market index. Hence, in an efficient market, plan sponsors would not rationally invest in (or pay active management fees for) an investment program which cannot outperform a market index. However, there exists a very large active pension fund management business in the United States. Our study begins to shed some light on the question of whether or not plan sponsors are behaving rationally to perpetuate this business. Our paper is organized as follows. Section I presents the models of selectivity and market timing used in this paper. Section II describes the data and methodology. Section III presents the empirical results. Section IV

presents a meta-analysis of our results. Section V discusses the results. Section VI concludes our paper.

## I. Models of Selectivity and Timing

It is important that portfolio managers be evaluated on both security selection ability and market timing skill. Furthermore, it has become standard practice to model selectivity and timing simultaneously. Jensen (1968, 1969) formulated a return-generating model to measure performance of the managed portfolios. The model is:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + u_{pt} \quad (1)$$

where  $R_{pt}$  is the excess (net of risk-free rate) return on the  $p$ th portfolio.  $R_{mt}$  is the excess (net of risk-free rate) return on the market portfolio,  $\alpha_p$  is a measure of security selection skill,  $\beta_p$  measures the sensitivity of the portfolio to the market return,  $u_{pt}$  is a random error which has expected value of zero and  $t$  denotes time. This specification assumes that the risk level of the portfolio under consideration is stationary through time and ignores the market timing skill of the managers. Indeed, portfolio managers may shift the overall risk composition of their portfolio in anticipation of broad market price movements. Fama (1972) and Jensen (1972) addressed this issue and suggested a somewhat finer breakdown of performance.

Treynor and Mazuy (1966) added a quadratic term to equation (1) to test for market timing ability. They argued that if a manager can forecast market returns, he will hold a greater proportion of the market portfolio when the return on the market is high and a smaller proportion when the return on the market is low. Thus, the portfolio return will be a nonlinear function of the market return as follows:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + \gamma (R_{mt})^2 + \epsilon_{pt} \quad (2)$$

A positive value of  $\gamma$  would imply positive market timing skill.

Jensen (1972) developed a similar model to detect selectivity and timing skill of managers. Jensen's measure of market timing performance calls for a fund manager to forecast the deviation of the market portfolio return from its consensus expected return. By assuming that the forecasted return and the actual return on the market have a joint normal distribution, Jensen shows that, under this assumption, a market timer's forecasting ability can be measured by the correlation between the market timer's forecast and the realized return on the market. He concluded that, under the above structure, the separate contributions of selectivity and timing cannot be identified unless, for each period, the manager's forecast and consensus expected return on the market portfolio,  $E(R_m)$ , are known.

Bhattacharya and Pfleiderer (1983) extended the work of Jensen (1972). By correcting an error made in Jensen (1972), they show that one can use a simple regression technique to obtain measures of timing and selection ability. Jensen assumes that the manager uses the unadjusted forecast of the market return in the timing decision. Bhattacharya and Pfleiderer assume that the manager adjusts forecasts to minimize the variance of the forecast error. They specify a relationship in terms of observable variables, which is similar to the Treynor and Mazuy model:

$$R_{pt} = \alpha_p + \Theta E(R_m)(1 - \Psi)R_{mt} + \Psi \Theta (R_{mt})^2 + \Theta \Psi \epsilon_{tR_{mt}} + u_{pt} \quad (3)$$

where

$\alpha_p$  = security selection skill (risk-adjusted excess return),

$\Theta$  = the fund manager's response to information; i.e., risk-level deviation from the

target risk-level depending on the optimal forecast of the market return,

$\Psi$  = the coefficient of determination between the manager's forecast and the excess return on the market, and

$\epsilon_t$  = the error of the manager's forecast

This quadratic regression of  $R_{pt}$  on  $R_{mt}$  allows us to detect the existence of stock selection ability as revealed by  $\alpha_p$ . The disturbance term in equation (3):

$$\omega_t = \Theta\Psi\epsilon_t R_{mt} + u_{pt} \quad (4)$$

contains the information needed to quantify the manager's timing ability. We can extract this information by regressing  $(\omega_t)^2$  on  $(R_{mt})^2$ :

$$(\omega_t)^2 = \Theta^2\Psi^2(\sigma_\epsilon)^2(R_{mt})^2 + \varsigma_t, \quad (5)$$

where

$$\varsigma_t = \Theta^2\Psi^2(R_{mt})^2[(\epsilon_t)^2 - (\sigma_\epsilon)^2] + (u_{pt})^2 + 2\Theta\Psi R_{mt}\epsilon_t u_{pt}. \quad (6)$$

The proposed regression produces a consistent estimator of  $\Theta^2\Psi^2\sigma_\epsilon^2$ , where  $(\sigma_\epsilon)^2$  is the variance of the manager's forecast error. Using the consistent estimator of  $\Theta\Psi$ , which we recover from equation (3), we obtain  $(\sigma_\epsilon)^2$ . This, coupled with knowledge about  $(\sigma_\pi)^2$ , the variance of excess return on the market, allows us to estimate  $\Psi = (\sigma_\pi)^2/[(\sigma_\pi)^2 + (\sigma_\epsilon)^2] = \rho^2$ , where  $\rho$  is the correlation between the manager's forecast and excess return on the market. Finally, we calculate  $\rho$  which truly measures the quality of the manager's timing information.

The Bhattacharya and Pfleiderer model of equation (3) is a refinement of the Treynor and Mazuy model. It focuses on the coefficient of the squared excess market return as an indication of timing skill. It is the first model that analyzes the error term to identify a manager's forecasting skill. Such a refinement should make the model more powerful than previous ones.

Further detail and econometric issues relating to the Bhattacharya and Pfleiderer model are discussed in Section II. In the empirical tests reported in Section III, we employed both the Treynor and Mazuy and the Bhattacharya and Pfleiderer models. This will allow us to examine the sensitivity of results to alternative model specifications.

There are other models in the literature that permit identification and separation of selectivity and timing skills of portfolio managers; i.e., models by Grinblatt and Titman (1989b), Henriksson and Merton (1981), and an alternative to the Henriksson and Merton model by Kon and Jen (1978, 1979). The Grinblatt and Titman model requires the historical sequence of portfolio weights (i.e., the amount invested in each stock) for the manager. Unfortunately, data on portfolio weights are very costly, time-consuming and not often available. The Henriksson and Merton model provides no significant advantage over the Bhattacharya and Pfleiderer model. The weakness of the Henriksson and Merton model is that it only tests whether the manager has special information, but it does not test whether the manager uses the information correctly (Dybvig and Ross (1985)). The forecasters in this model are less sophisticated than those of the Bhattacharya and Pfleiderer model, where they do forecast how much better the superior investment will perform. Henriksson and Merton assume that managers have a coarse information structure in which dichotomous signals are only predictive of the sign of the excess return of the market relative to the risk-free rate. In their model, the probability of receiving an "up" or a "down" signal in no way depends upon how far the market will be "up" or "down."

## **II. Data and Methodology**

The data for this study consist of monthly returns for the period January 1983 through



December 1990 (96 months) for a random sample of 71 U.S. equity pension fund managers with complete data for the entire period. The 71 managers were chosen from the Russell pension manager database by random drawing (with replacement) so as to reflect as close as possible the actual distribution of managers (by investment style) in the database. Returns are net of expenses and management fees. The data include returns on tax-free, fully discretionary equity portfolios that are at least \$5 million in size. These portfolios are managed by banks, insurance companies and investment advisors who have been allocated funds by pension plan sponsors. The identities of the fund managers and sponsors are not included. The managers invest exclusively in the U.S. equity market. The random sample of pension fund managers was provided by the Frank Russell Company of Tacoma, Washington. Among other services, the Frank Russell Company evaluates the performance of the managers of a number of pension funds throughout the United States. The Frank Russell Company segregates equity managers into four basic investment styles on the basis of managers' portfolio characteristics. These are: (1) Earnings Growth, (2) Market-Oriented, (3) Price-Driven, and (4) Small Capitalization. Our sample consists of 18 Earnings Growth, 19 Market-Oriented, 18 Price-Driven, and 16 Small Capitalization managers. Appendix I.A describes these four investment styles. Monthly returns on the 91-day Treasury bill were used as a proxy for the risk-free rate.

Our study uses several alternative equity benchmark portfolios. Two of these are the S&P 500 Index and the Russell 3000 Index. The Russell 3000 Index is a broad equity market index like the S&P 500. Appendix I.B describes the Russell 3000 Index and compares it to the S&P 500 Index. The idea of investment "style management" is becoming increasingly important to both academic studies and professional investment management (see, e.g., Tierney and

Winston (1991)). Therefore, in addition to the two broad equity market indices, we also use four style indices as benchmarks. To be more specific, we use separate benchmarks for four different investment styles. These style indices are the Russell 1000 Index (for Market-Oriented managers), the Russell 2000 Index (for Small Cap managers), the Russell Price-Driven Index (for Price-Driven managers), and the Russell Earnings Growth Index (for Earnings Growth managers). Appendix I.B describes these indices and compares them to broad market indices. The use of several alternative indices will allow us to examine the sensitivity of pension fund manager's performance to alternative benchmarks. An estimate of the variance of the excess return on the market,  $(\sigma_{\epsilon})^2$ , was derived from observed returns for each benchmark following the procedure of Lee and Rahman (1990).

In the empirical test, it is necessary to correct for heteroscedasticity in both the Treynor and Mazuy model and the Bhattacharya and Pfleiderer model. In the Treynor and Mazuy model, the error term will exhibit conditional heteroscedasticity because of the fund manager's attempt to time the market, even though security returns are assumed to be independent and identically distributed through time. To correct this, following Breen, Jagannathan and Ofer (1986) and Lehmann and Modest (1987), we use heteroscedasticity-consistent standard errors proposed by White (1980), Hansen (1982), and Hsieh (1983). The significance tests reported in Section III are based on heteroscedasticity-adjusted t-statistics.

In the Bhattacharya and Pfleiderer model, the procedure discussed in Section I does not produce the most efficient estimates of the parameters since the disturbance term in equations (3) and (5) are heteroscedastic. More efficient estimates can be obtained by taking into account the heteroscedasticity of the disturbance terms. We followed a Generalized Least Squares (GLS)

procedure, which makes a correction for heteroscedasticity, to obtain efficient estimates of parameters. This methodology is more fully described in Lee and Rahman (1990).

As noted in Coggin and Hunter (1993), one weakness of the Treynor and Mazuy and the Bhattacharya and Pfliederer models is that they ignore negative or inferior market timing. We modify these models to allow negative timing skill. We hypothesize that managers can exhibit negative ex post timing skill. In the Treynor and Mazuy model, this means the managers hold a smaller portion of the market portfolio when the market return is high. In the Bhattacharya and Pfliederer model, this is indicative of a negative correlation between the manager's beta and the market return. Such results in both models could be due to the inability of managers to correctly forecast the expected return on the market portfolio. Hence these managers would forecast the market return to be high when it is actually low and vice versa. In the Treynor and Mazuy model of equation (2), a negative value of  $\gamma$  would be indicative of poor market timing.

For the Bhattacharya and Pfliederer model, we examine the sign of the coefficient of  $(R_m)^2$  in equation (3). Intuitively, in the spirit of the Treynor and Mazuy model, the sign of this coefficient will be indicative of the nature of timing skill. If the estimated value of this coefficient is negative, we designate timing skill (given by  $\rho$ ) to be poor (negative). This modification makes these models more realistic. A similar adjustment of the Bhattacharya and Pfliederer model was implicitly introduced in Jagannathan and Korajczyk (1986, p. 229).

### **III. Empirical Results**

#### **A. Significance Tests for Performance Measures**

Table I presents summary results from the two models. In this table and those that

follow, "S&P 500" denotes results based on using the S&P 500 as the benchmark portfolio, "Russell 3000" denotes results based on using the Russell 3000 as the benchmark portfolio, and "Style Index" denotes results based on using each manager's appropriate style index as the benchmark portfolio. These results show some evidence of positive security selection skill and negative timing skill on the part of managers. The number of significant positive selectivity values exceeds the number of significant negative selectivity values for both models regardless of the benchmark used. For timing skill, the results are just the opposite. For both models, the number of significant negative timing values exceeds the number of significant positive timing values regardless of the benchmark used.

---- Insert Table I about here ----

## **B. Mean Values of Performance Measures**

Table II presents the means of the selectivity and timing values for all managers and for the subsets of managers classified by investment style. For the entire sample (All Managers), both models show a positive mean selectivity value for all three alternative benchmarks. These values are significant at the .05 level for two of the three benchmarks. For timing skill, the results are just the opposite. For the entire sample, both models show a negative mean timing value for all three alternative benchmarks. However, for only one of the three benchmarks (the S&P 500), the mean timing value is significant at the .05 level for both models. Hence the results using the S&P 500 Index as a benchmark contrast with the results obtained using the Russell 3000 Index and the Style Indices as benchmarks. As shown in Appendix I.B, the latter two indices are much more representative of the managers' investment universe (i.e., true

investment opportunities) than the former and, as such, are more appropriate benchmarks than the former.

The results in Tables I and II suggest that pension fund managers are on average better stock pickers than market timers. The results that were only hinted at in Table I are now strongly supported in Table II. Our results relating to selection skill are consistent with those of Lee and Rahman (1990), who found some evidence of superior selection skill on the part of mutual fund managers. They also found evidence of superior market timing skill for several managers. However, it should be pointed out that Lee and Rahman (1990) ignored negative market timing skill in their model, while we allow negative market timing here. Our market timing results are consistent with those of previous studies on mutual fund performance (see Kon (1983), Chang and Lewellen (1984), Henriksson (1984), Lehmann and Modest (1988), Cumby and Glen (1990), and Connor and Korajczyk (1991)). These studies found more evidence of negative market timing than positive. These studies also found some evidence of negative selection skill for mutual funds.

There are differences in the portfolio characteristics and investment styles among the Earnings Growth, Market-Oriented, Price-Driven, and Small Capitalization managers. It is therefore useful to examine performance measures for each investment style separately. Table II presents mean values of the performance measures for each style of manager. It also provides the aggregated rank of each group. These ranks do not vary between the models for a given benchmark. However, they do vary somewhat across benchmarks for a given model.

The period 1983-1990 was a period in which the overall stock market was up substantially. For the eight years, the Russell 3000 grew at an annualized rate of 14.17%, and

the S&P 500 grew at a 15.60% rate. For the majority of this period (up until the end of 1988) the "value" investment style was favored by the market relative to other investment styles. Our benchmark for this style is the Price-Driven index which grew at an annualized rate of 15.53%. This compares to the "growth" investment style (represented by the Earnings Growth index) which grew at a 13.72% rate, and the Small Capitalization style (represented by the Russell 2000 index) which grew at a 7.38% rate. In Table II we see that, using the broad stock market indices as benchmarks, a negative mean selectivity value is consistently observed for the Earnings Growth and Small Capitalization managers. This is consistent with the preference of the stock market for the period. However, if we look at the Style Index as a benchmark, we see that these managers (as well as all other styles) have positive selectivity values. Thus, while we observe a positive mean selectivity value and a negative mean timing value across All Managers for each benchmark, it does begin to matter which benchmark portfolio is used (and, perhaps, which time period) when we examine the results for the four investment styles.

Finally, one needs to be somewhat concerned about the size of the timing values. At a purely statistical level, one can assess the significance of the timing values by looking at the t-tests. However, in the Treynor-Mazuy model the impact of timing on portfolio return is, in effect, measured by multiplying a rather small decimal fraction,  $\gamma$ , by a squared decimal fraction,  $(R_{mt})^2$ . Thus, at the level of actual portfolio returns, there is a relatively small reward/penalty to this activity in our data. Further research in the area of the measurement and assessment of market timing would help clarify this issue.

---- Insert Table II about here ----

### C. Correlation of Performance Measures

To examine the sensitivity of results to benchmarks and models, we also examine the correlation of the performance measures across models and benchmarks. We use two measures of association - the Pearson correlation coefficient and the Spearman rank correlation coefficient. Tables III, IV, and V provide correlational summaries of the results presented in Tables I and II. Table III represents the correlation of a performance measure (selectivity or timing) with itself between benchmarks for a given model. All the correlations reported in the table are significant at the .0001 level. There is a very high correlation between the results based on the broad market indices - the S&P 500 Index and the Russell 3000 Index. The performance measures based on these benchmarks are somewhat less correlated with those based on style indices. These results are consistent for both models and for both the selectivity and timing measures. Table IV presents the correlation of a performance measure (selectivity or timing) with itself between models for a given benchmark. These correlations are very high and significant at the .0001 level. The results in Tables III and IV indicate high ranking consistency among benchmarks and between models.

Finally, we present the correlation between selectivity and timing skill within a model for a given benchmark. These correlations are given in Table V. All these correlations are significantly negative. We will have more to say about this result in Section V.B.

---- Insert Tables III, IV, V about here ----

### IV. Meta-Analysis of Results

Meta-analysis is a parametric statistical technique for the cumulation of results across studies

or units of analysis. The contribution of meta-analysis is to offer a statistical technique to produce direct estimates of the mean and standard deviation of population values. Thus meta-analysis allows more statistically powerful inferences from data than are possible using more traditional disaggregated analyses. Following its early beginnings in physics and psychology, meta-analysis has recently been applied to cumulate results across studies in several other disciplines including accounting (Christie (1990) and Trotman and Wood (1991)), finance (Coggin and Hunter (1983, 1987, 1993) and Dimson and Marsh (1984)), and marketing (Farley and Lehman (1986)). Recent comprehensive texts on meta-analysis include Hedges and Olkin (1985) and Hunter and Schmidt (1990).

There are a number of "study artifacts" which can cause the results of one study to appear different or even contradictory to those of another. Among the more prominent artifacts are sampling error, error of measurement, and restriction of range on the dependent variable. These artifacts are discussed in detail in Hunter and Schmidt (1990, Chapters 2 and 3). In this paper, we focus on sampling error in the regression values for selectivity and market timing across managers. Meta-analysis has been primarily developed for correlational data. However, the time series regressions performed in our paper have identical specifications (by performance measurement model) across the sample of pension fund managers. Thus, for the purpose of meta-analysis, we can consider each of the 71 managers as a "study," cumulate the results and apply meta-analysis. Appendix II to this paper presents a brief discussion of the meta-analysis technique for regression coefficients used in this section.

As discussed in Appendix II, the standard meta-analysis formulas must be adjusted for the effect of correlated regression residuals. Table VI presents the average correlations between



the regression residuals for the 71 managers for the entire sample period across benchmarks and models. These average correlations were used in the adjusted formulas to calculate the meta-analysis results given in Tables VII and VIII.

— Insert Table VI about here —

Table VII presents the results of the meta-analysis of the selectivity and timing coefficients based on three benchmark portfolios and using heteroscedasticity corrected t-values. The first row of this table gives the frequency-weighted mean of the observed values for each parameter,  $\bar{b}$ ; the second row gives estimates of the standard deviation of the observed values,  $s_b$ ; the third row gives estimates of the standard deviation of the population values,  $s_\beta$ ; the fourth row gives estimates of the frequency-weighted average squared deviation of the observed values,  $s_b^2$ ; the fifth row gives estimates of the variance of the population values,  $s_\beta^2$ ; the sixth row gives estimates of the sampling error variance,  $s_e^2$ ; the seventh row gives the chi-square value for the ratio of the observed variance to the sampling error variance; and the last row gives estimates of the proportion of total observed variance accounted for by sampling error,  $s_e^2/s_b^2$ .

--- Insert Table VII about here ---

#### A. Selectivity

For selectivity, the mean monthly values in Table VII are positive in every case but very small. However, on an annualized basis, these numbers become more meaningful. For the Bhattacharya and Pfleiderer model, the annualized mean selectivity values are .41% (S&P 500), .93% (Russell 3000), and 1.97% (Style Index). For the Treynor and Mazuy model, the annualized mean selectivity values are .51% (S&P 500), .96% (Russell 3000), and 1.99% (Style

Index). Hence we see that for both models, managers do better on average relative to their own style index as compared to the broader market indices. This result is instructive, since much of the common investment wisdom implies that investment managers "can't beat the market." This result suggests that such a comment begs an important question regarding which benchmark should be used in evaluating a manager. We remind the reader that these returns are net of investment management fees.

The chi-square values are significant at the .05 level or less for the selectivity values using all three benchmarks for both models. This implies that there is real variation (in excess of that attributable to sampling error) around the mean selectivity value in each case.

## **B. Timing**

For market timing, the mean values in Table VII are negative in each case. This result is consistent with the results of Kon (1983), Chang and Lewellen (1984), Henriksson (1984), Grinblatt and Titman (1988), Lehmann and Modest (1988), Cumby and Glen (1990), Coggins and Hunter (1993), and Connor and Korajczyk (1991) who examined mutual fund returns. Furthermore, the chi-square values are significant at the .05 level or less in each case. Thus in every case there is evidence of real variation around the negative mean timing value.

If there were no real variation around the observed mean value, then the observed mean would be the true value for each of the 71 managers. However, in our case, there is evidence of real variation in every set of selectivity and market timing values. To put these results in perspective, we can look at the last row of Table VII for each model and examine the proportion of total observed variance accounted for by sampling error. For the Bhattacharya and Pfleiderer

model, the percentage of observed variance in selectivity accounted for by sampling error goes from 57% to 57% to 50% across benchmarks; while the percentage of variance in timing accounted for by sampling error goes from 73% to 74% to 63% across benchmarks. For the Treynor and Mazuy model, the percentages for selectivity go from 57% to 57% to 50% across benchmarks; while the timing percentages go from 14% to 15% to 13% across benchmarks. We should note that, as discussed in Hunter and Schmidt (1990), these percentages of variance attributable to sampling error may well contain other unaccounted for study artifacts (such as measurement error).

### **C. The 80% Probability Intervals for Selectivity and Timing**

Assuming selectivity and market timing to be normally distributed, we can also examine the 80% probability intervals (i.e., the lower and upper 90% probability values) for the spread of the observed and population values presented in Table VIII. The probability intervals in Table VIII clearly show the amount of variation in both the observed and the population values for selectivity and market timing. As noted above, there is real variation in selectivity and timing values in every case. The 80% probability intervals for selectivity are all shifted towards positive values, while the 80% probability intervals for timing are all shifted towards negative values. This result is confirmed by the significance counts for positive and negative selectivity and timing values in Table I.

Using the 80% probability intervals for the population selectivity values in Table VIII, we can look at the true spread in pension manager excess returns for the two models across benchmarks. The return for the top 10% of managers is obtained by annualizing the appropriate

upper bound return in Table VIII, and the return for the bottom 10% of managers is obtained by annualizing the appropriate lower bound return in Table VIII. For the Bhattacharya and Pfleiderer model using the S&P 500 benchmark, the true annualized spread in returns is 5.63% (top 10% = 3.26%, bottom 10% = -2.37%); using the Russell 3000, the true spread is 5.79% (top 10% = 3.86%, bottom 10% = -1.93); and using the style index, the true spread is 5.65% (top 10% = 4.83%, bottom 10% = -.82%). For the Tryenor and Mazuy model, the true annualized spread in returns using the S&P 500 benchmark is 5.93% (top 10% = 3.51%, bottom 10% = -2.42%); using the Russell 3000, the true spread is 5.84% (top 10% = 3.92%, bottom 10% = -1.92%); and using the style index, the true spread is 6.03% (top 10% = 5.05%, bottom 10% = -.98%). Hence there is evidence in our data that the best pension fund managers can deliver substantial risk-adjusted excess returns, no matter which model or benchmark we use. This complements the results of Grinblatt and Titman (1989a), Ippolito (1989), Lee and Rahman (1990), and Coggin and Hunter (1993) who found some evidence of superior performance in their studies of mutual funds.

-- Insert Table VIII about here --

## V. Discussion

### A. Sensitivity of Results to Benchmarks and Models

Our general finding is that selectivity is positive and timing is negative on average across all models and benchmarks. The results in Tables III and IV indicate that the overall rankings of both performance measures are not very sensitive to alternative benchmarks and models in our data. However, we did observe some sensitivity of results to the choice of a benchmark

when we divided up the managers by investment style in Table II.

Our basic results differ from those of Lehmann and Modest (1987) and Grinblatt and Titman (1989a), who found that performance results varied across models and benchmark portfolios. It should be pointed out that there is a problem in the Lehmann and Modest (1987) analysis. They examined selectivity in the context of a Jensen-like measure using the CAPM and APT models. Market timing and factor timing activities are not included in their analysis. Market timing was also ignored by Grinblatt and Titman (1989a). Grant (1977) explained how market timing actions will affect the results of empirical tests that focus only on selection skill. He showed that market timing ability will cause the observed regression estimate of selectivity to be downwardly biased. The results of Lee and Rahman (1990) are consistent with Grant's (1977) contention. A similar conclusion was drawn by Chang and Lewellen (1984) and Henriksson (1984). Moreover, as Jensen (1972), Admati and Ross (1985), Dybvig and Ross (1985), and Grinblatt and Titman (1989b) have shown, the Jensen-like measure may penalize the performance of market timers.

## **B. Negative Correlation Between Selectivity and Timing**

As discussed in Sections III and IV, we observe a strong negative correlation between selectivity and market timing in our data. Furthermore, this is consistent with the results of several other studies. The literature on investment management contains a number of studies documenting the negative market timing ability of mutual fund managers (see Chua and Woodward (1986) for a summary and extension of these studies). Ours is the first study we know of which documents this finding for pension fund managers.

The negative correlation between selectivity and timing presents a problem of interpretation. Hunter, Coggin and Rahman (1992) show that, for the regression models used in this study, the correlation between the estimates of selectivity and timing will necessarily be negative. They show that this is because the sampling errors for the two estimates are negatively correlated. The magnitude of the negative correlation between the two estimates is the same as the magnitude of the negative correlation between the two sampling errors. Thus, once we account for the effect of negatively correlated sampling errors, selectivity and timing are largely uncorrelated in our data.

The correlation between selectivity and market timing is currently an unsettled question in the literature. Coggin and Hunter (1993) calculated a corrected correlation of about  $-.64$  in the Lee and Rahman (1990) data, but noted that this was also an artifact of correlated sampling errors for the two estimates. They also calculated an observed correlation of  $.04$  ( $N=37$  mutual funds) between "Overall Selectivity" and "Overall Timing" in Kon (1983, Table 5), who used a different model of market timing. Using the Henriksson and Merton (1981) model, Henriksson (1984) and Connor and Korajczyk (1991) report a negative correlation. Lehmann and Modest (1987, fn. 33) report basically no "substantive correlation" between the two. Jagannathan and Korajczyk (1986) have presented an argument that the observed negative correlation between selectivity and timing could also be the result of some other phenomenon, such as changes in firms' debt/equity ratios in relationship to that of the benchmark portfolio. Finally, Grinblatt and Titman (1989b) have shown that many of the desirable properties of a performance measurement model which seeks to estimate both selectivity and market timing skill are not present if selectivity and timing are correlated.

### **C. Survivorship Bias**

The issue of survivorship bias is well known in studies of investment performance. A recent study by Brown, Goetzmann, Ibbotson and Ross (1992) highlights this issue with regard to performance measurement. The basic issue here is as follows. Our study includes 71 pension managers with complete data from 1983 to 1990. Hence, any manager who may have disappeared through merger or poor performance is not included in our data. To the extent that our sample underrepresents such managers, our results are biased in favor of more successful managers. We do not know the true extent of this bias in our results, but the results in Grinblatt and Titman (1989a) suggest that it is not large.

## **VI. Summary and Conclusion**

This paper presents an empirical investigation of the security selection and market timing performance of a random sample of 71 U.S. equity pension fund managers for the period 1983-1990. Our major findings are as follows. Regardless of the choice of benchmark portfolio or estimation model, the selectivity measure is positive on average and the timing measure is negative on average. However, both selectivity and timing do appear to be somewhat more sensitive to the choice of a benchmark (and, possibly, the time period) when managers are classified by investment style. A meta-analysis of the regression results was performed to quantify the effect of sampling error. In every case, meta-analysis revealed some real variation (in excess of that attributable to sampling error) around the mean values for each measure. An examination of the 80% probability intervals for selectivity revealed that the best equity pension

fund managers can deliver substantial risk-adjusted excess returns. Consistent with previous studies of equity mutual fund performance, we also found a negative correlation between selectivity and timing. However, we argue that the observed negative correlation in our data is largely an artifact of negatively correlated sampling errors for the two estimates.

Much work remains to be done in this area. While active equity managers are currently losing ground to passively managed index funds, actively managed equities still represent the largest fraction of the equity component of corporate pension funds. We still do not know why some active managers are able to provide substantial risk-adjusted performance, while most cannot. Identifying the characteristics of successful money managers should be one focus of future research. Furthermore, while there are some interesting statistical explanations, we still do not have a satisfactory substantive model of the relationship between the security selection and market timing ability of active equity managers. This is another fertile area for study.



## Appendix I

This Appendix is based on Haughton and Christopherson (1989).

### A. Style Descriptions

1. **Earnings Growth:** Earnings Growth managers focus predominantly on earnings and revenue growth and attempt to identify companies with above-average growth prospects. In general, two basic categories of securities are owned by Earnings Growth managers: (a) companies with consistent above-average (historical and prospective) profitability and growth, and (b) companies expected to generate above-average near-term earnings momentum based upon company, industry, or economic factors.
2. **Market-Oriented:** Market-Oriented managers are broadly diversified managers who participate in all sectors of the equity market. The portfolios of these managers may either be well diversified, or take meaningful sector/factor bets relative to the market toward both growth and value over time. Market-Oriented managers are typically willing to consider companies representative of the broad market when seeking investment opportunities.
3. **Price-Driven:** Price-Driven managers focus on the price and value characteristics of a security in the selection process. These managers buy stocks from the low price portion of the market, and are sometimes called value or defensive/yield managers. In general, these managers focus on securities with low valuations relative to the broad market.

4. Small Capitalization: Small Capitalization managers focus on small capitalization stocks. These companies may be unseasoned and rapidly growing but sometimes are simply small businesses with long histories. Typical characteristics of small capitalization portfolios are below-market dividend yields, above-market betas, and high residual risk relative to broad market indices.

B. Description of Russell Indices

(Note: All Russell indices are capitalization-weighted)

Benchmarks for Aggregate Portfolios

Russell 3000 Index: The Russell 3000 Index includes the top 3000 U.S. companies ranked by capitalization. Haughton and Christopherson (1989) discussed two reasons for choosing the Russell 3000 Index over the S&P 500 Index.

- (1) The S&P 500 spans only 75% of the investable U.S. equity market. As such, it has a large capitalization bias but, within large cap stocks, it excludes some large companies. It also includes non-U.S. companies, so it is not strictly a U.S. equity market benchmark. There is no adjustment in the index for cross-ownership of shares, resulting in the overweighting of certain companies. Since it covers only 500 companies, it does not reflect many of the long-term bets managers make away from the index.

- (2) The Russell 3000 covers 98% of the investable U.S. equity market. It weights all market sectors according to their investment opportunities, and is confined to U.S. companies and hence has no foreign exposure. It is adjusted for cross-ownership, thereby reflecting true investment opportunities; and spans nearly all of the stocks in which a manager is likely to invest. Hence, the index is relatively unbiased.

### Style Indices

Broad equity market benchmarks like the S&P 500 and the Russell 3000 are suitable for evaluating pension managers who use the whole market as a base. Many U.S. equity pension managers specialize in subsets of the market. As such, a finer set of performance benchmarks that more closely match the investment styles of individual managers is needed to ensure identification of elements attributable to investment styles. The Frank Russell Company maintains four style indices - one for each investment style. The key fundamental characteristics of each style index are similar to the equity profile of a typical manager of that style. This indicates that the subuniverse of stocks that comprise the style indices contains the type of stocks from which each style of managers would normally choose; i.e., they constitute rough "normal" portfolios. These style benchmarks are much more representative of the specialized managers' selection universes than the broad market and hence should provide better tools for performance evaluation. These style indices are:

1. Russell 1000 Index: The Russell 1000 is the benchmark recommended for Market-Oriented style managers. It is composed of the top 1000 stocks in the Russell 3000 Index ranked by capitalization. Hence, it focuses on the broad-based large cap segment of the market and encompasses about 90% of all the equity opportunities in the U.S. equity market.
2. Russell 2000 Index: The Russell 2000 is the small cap benchmark and is useful for evaluating small capitalization style managers. It is composed of the smallest 2000 stocks in the Russell 3000 Index ranked by capitalization. Of the 10% of the total U.S. equity market comprised of small stocks, the Russell 2000 Index covers about 8%.
3. Earnings Growth Index: Earnings Growth Index is an index for Earnings Growth style managers, and is composed of those securities in the Russell 1000 Index that have above-average growth prospects.
4. Price Driven Index: Price Driven Index is an index for Price Driven style managers. It is composed of those securities in the Russell 1000 Index that have low valuations relative to the broad market. "Low valuation" is defined by examining financial ratios such as the P/E ratio, dividend yield, the price/book ratio, and the price/sales ratio.

## Appendix II

### The Meta-Analysis of Regression Values

#### A. Theoretical Meta-Analysis Parameters

This Appendix is taken from a more detailed presentation given in Coggin and Hunter (1993). Meta-analysis was developed as a parametric statistical technique to cumulate results across studies or units of analysis. In this appendix, we will use the words "study," "manager," and "portfolio" interchangeably. We initially assume that the number of managers to be analyzed is large enough that we can ignore sampling error due to a finite number of managers, and concentrate on sampling error in regression estimates for individual managers. We also assume that the specification of each regression equation is identical across managers. We denote observed regression values (including the intercept) as  $b$ , population values as  $\beta$ , and sampling error as  $e$ . Thus:

$$e = b - \beta \text{ or } b = \beta + e \quad (\text{A-1})$$

The average observed value is:

$$\bar{b} = \bar{\beta} + \bar{e} \quad (\text{A-2})$$

Across a large number of managers, the average error,  $\bar{e}$ , will be zero; thus  $\bar{b} = \bar{\beta}$ .

Since we are comparing the portfolios of pension fund managers, we denote each manager by the subscript  $i$ . Then:

$$b_i = \beta_i + e_i \quad (\text{A-3})$$

Across portfolios,  $\beta$  and  $e$  will be uncorrelated, so that the variance of observed values,  $\sigma_b^2$ , will be larger than the variance of population values,  $\sigma_\beta^2$ , by the amount of sampling error,  $\sigma_e^2$ :

$$\sigma_b^2 = \sigma_\beta^2 + \sigma_e^2 \quad (\text{A-4})$$

From equation (A-4), the variance of the population regression values can be written as:

$$\sigma_\beta^2 = \sigma_b^2 - \sigma_e^2 \quad (\text{A-5})$$

The key to meta-analysis is the fact that the sampling error variance,  $\sigma_e^2$ , can be computed using known statistical theory. Thus equation (A-5) becomes a formula to compute the population variance,  $\sigma_\beta^2$ .

## B. Estimating Meta-Analysis Parameters

In the previous section, we assumed that the number of studies to be cumulated is large. Specifically, this implies that the observed variance of the sampling errors would equal the theoretical sampling error variance. If the number of studies is small, then the observed variance of the sampling errors will differ by chance from the theoretical sampling error variance. Hence we use the notation "s<sup>2</sup>" for the estimated variances below.

If a population value is assumed be constant across studies, Hunter and Schmidt (1990) show that the best estimate of that value is its frequency-weighted average:

$$\bar{b} = \Sigma[N_i b_i] / \Sigma N_i \quad (\text{A-6})$$

where  $b_i$  is the observed value in study  $i$  and  $N_i$  is the number of observations in study  $i$ . The corresponding observed variance estimate across studies is the frequency-weighted average squared deviation:

$$s_b^2 = \Sigma[N_i(b_i - \bar{b})^2] / \Sigma N_i \quad (\text{A-7})$$

The observed variance estimate,  $s_b^2$ , is a confounding of two sources of variation: variation in population values (if any) and variation in observed values due to sampling error. Thus an

estimate of the variation in population values can only be obtained by correcting the observed variance estimate,  $s_b^2$ , for sampling error. Hunter and Schmidt (1990) show that sampling error across studies behaves like error of measurement, and the resulting formulas are comparable to the standard formulas in classic psychometric measurement theory or reliability theory.

From classic psychometric theory (Thorndike (1982)), we have:

$$\text{Observed value} = \text{true value} + \text{error of measurement} \quad (\text{A-8})$$

where the true value and error of measurement are uncorrelated. Hence:

$$\text{Observed variance} = \text{true variance} + \text{error variance} \quad (\text{A-9})$$

In meta-analysis, it is similarly true that the population regression values,  $\beta_i$ , and the sampling error,  $e_i$ , are uncorrelated across studies. Therefore we can write:

$$\text{Observed variance} = \text{population variance} + \text{sampling error variance} \quad (\text{A-10})$$

$$s_b^2 = s_\beta^2 + s_e^2 \quad (\text{A-11})$$

The observed variance estimate,  $s_b^2$ , is the frequency-weighted average squared deviation defined above. The sampling error variance estimate required by meta-analysis is then:

$$s_e^2 = \Sigma[N_i(\text{standard error } b_i)^2] / \Sigma N_i \quad (\text{A-12})$$

The population variance (sometimes called the "corrected variance") can thus be estimated as:

$$s_\beta^2 = s_b^2 - s_e^2 \quad (\text{A-13})$$

Equation (A-13) is the fundamental estimating equation for the theoretical values in equation (A-5).

Consider a special case where  $\sigma_\beta^2=0$ ; i.e., the homogeneous case where  $\beta$  has the same value across studies and  $\sigma_b^2=\sigma_e^2$ . The observed variance,  $s_b^2$ , has a sampling error variance that

depends on the number of studies. With probability approximately .50,  $s_b^2$  will be less than  $\sigma_b^2$ ; and hence less than  $\sigma_e^2$ . For such a sample of studies, the estimated variance of  $\beta$  will be negative (as in ANOVA when the sample F value is less than one). If the variance estimate is negative, it means that the best estimate of the population variance is zero. In a statistical model where variance is estimated by subtraction (as in ANOVA), a negative value does not imply that the model is inconsistent or that the computations are incorrect.

The estimate of the population variance,  $s_b^2$ , can thus be positive, negative or zero. If it is negative or zero, the inference is that there is no variation in observed values that cannot be attributed to sampling error. That is, all variance in observed values is artifactual. If the corrected variance across studies is positive, it may still be trivial in size.

#### C. A Significance Test for Real Variation Across Studies

The hypothesis that there is no real variation in observed values has a statistical test. The ratio of the observed variance estimate to the sampling error variance estimate has a chi-square distribution with  $k-1$  degrees of freedom:

$$\chi^2 = ks_b^2/s_e^2 \quad (A-14)$$

where  $k$ =number of studies.

This statistic can be used as a formal test of no variation; although if  $k$  is large, it has high statistical power and may reject the null hypothesis given even a trivial amount of real variation (Hedges and Olkin (1985), Cohen (1988), and Hunter and Schmidt (1990)). Thus if the chi-square value is not significant, there is strong evidence that there is no real variation across studies. However, if the  $k$  studies are not independent, then the power of the chi-square test is



reduced as discussed in the next section.

#### D. Independence

Given a set of regression estimates, there is a corresponding set of sampling errors. In the preceding discussion, it was assumed that the variance of the sampling errors across studies would itself differ only by sampling error from the hypothetical error variance across independent replications. This is true for most applications of meta-analysis and follows immediately from the independence of the estimates across studies. However, this is not always true.

In this study the impact of the market proxy is controlled. However, the portfolios of two equity pension fund managers may overlap. Hence the securities the two portfolios have in common will contribute their particular returns to both portfolio return sequences. The residuals of those securities will thus contribute to the residuals of the two portfolios. This means that the two portfolios will have time series regression residuals that are not entirely independent. It can be shown that the sampling errors for the two portfolio regressions will also be nonindependent and positively correlated. The strength of the correlation largely depends on the degree of fund overlap, and possibly on additional unmeasured common factors.

Consider the set of regression residuals for two portfolios. If the correlation between residuals is  $r$ , then the error variance across portfolios will not be  $s_e^2$ , but rather the product  $[(1-r)s_e^2]$ . The adjusted formulas for meta-analysis are therefore:

$$s_b^2 = s_\beta^2 + (1-r)s_e^2 \quad (A-15)$$

$$s_\beta^2 = s_b^2 - (1-r)s_e^2 \quad (A-16)$$

$$s_{\beta}^2 = (s_b^2 - s_e^2) + r s_e^2 \quad (\text{A-17})$$

Thus, standard meta-analysis formulas will underestimate the variance of  $\beta$ . In particular, the estimated variances for selectivity and timing will be too low by the amount  $rs_e^2$ . The adjusted formula for chi-square would thus be:

$$\chi^2 = k s_b^2 / (1-r) s_e^2 \quad (\text{A-18})$$

$$\chi^2 = [1/(1-r)](k s_b^2 / s_e^2) \quad (\text{A-19})$$

Hence, the standard chi-square test statistic for homogeneity of regression values given earlier in equation (A-14) would be an underestimate and thus would have somewhat lower than optimal power to detect departures from homogeneity. Therefore the standard chi-square test would be a conservative test for heterogeneity. All meta-analysis results presented in this study use the adjusted formulas for the variances and the chi-square test.

## REFERENCES

- Admati, A. and S.A. Ross, 1985, Measuring Investment Performance in a Rational Expectations Equilibrium Model, Journal of Business 58, 1-26.
- Berkowitz, S.A., L.D. Finney, and D.E. Logue, 1988, The Investment Performance of Corporate Pension Plans (Quorum Books, New York, NY).
- Bhattacharya, S. and P. Pfleiderer, 1983, A Note on Performance Evaluation, Technical Report 714, Stanford University, Stanford, CA.
- Breen, W., R. Jagannathan, and A.R. Ofer, 1986, Correcting for Heteroscedasticity in Tests for Market Timing Ability, Journal of Business 59, 585-598.
- Brinson, G.P., L.R. Hood, and G.L. Beebower, 1986, Determinants of Portfolio Performance, Financial Analysts Journal 42, 39-44.
- Brown, S.J., W.N. Goetzman, R.G. Ibbotson, and S.A. Ross, 1992, Survivorship Bias in Performance Studies, forthcoming in Review of Financial Studies.
- Chang, E.C. and W.G. Lewellen, 1984, Market Timing and Mutual Fund Investment Performance, Journal of Business 57, 57-72.
- Christie, A.A., 1990, Aggregation of Test Statistics, Journal of Accounting and Economics 12, 15-35.
- Chua, J.H. and R.S. Woodward, 1986, Gains from Market Timing, Monograph No. 1986-2, New York University, New York, NY.
- Coggin, T.D. and J.E. Hunter, 1983, Problems in Measuring the Quality of Investment Information: The Perils of the Information Coefficient, Financial Analysts Journal 39,

25-33.

- , and J.E. Hunter, 1987, A Meta-Analysis of Pricing "Risk" Factors in APT, Journal of Portfolio Management 14, 35-38.
- , and J.E. Hunter, 1993, A Meta-Analysis of Mutual Fund Performance, forthcoming in Review of Quantitative Finance and Accounting.
- Cohen, J., 1988, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. (Lawrence Erlbaum, Hillsdale, NJ).
- Connor, G. and R. Korajczyk, 1991, The Attributes, Behavior, and Performance of U.S. Mutual Funds, Review of Quantitative Finance and Accounting 1, 5-26.
- Cumby, R.E. and J.D. Glen, 1990, Evaluating the Performance of International Mutual Funds, Journal of Finance 45, 497-521.
- Dimson, E. and P. Marsh, 1984, An Analysis of Stock Brokers' and Analysts' Unpublished Forecasts of UK Stock Returns, Journal of Finance 39, 1257-1292.
- Dybvig, P. and S.A. Ross, 1985, Differential Information and Performance Measurement Using a Security Market Line, Journal of Finance 40, 383-399.
- Fama, E.F., 1972, Components of Investment Performance, Journal of Finance 27, 551-567.
- Farley, J.U. and D.R. Lehmann, 1986, Meta-Analysis in Marketing (Lexington Books, Lexington, MA).
- Grant, D., 1977, Portfolio Performance and the Cost of Timing Decisions, Journal of Finance 32, 837-846.
- Grinblatt, M. and S. Titman, 1988, The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns, Working Paper, University of California, Los Angeles, CA.

- and S. Titman, 1989a, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, Journal of Business 62, 393-416.
- and S. Titman, 1989b, Portfolio Performance Evaluation: Old Issues and New Insights, Review of Financial Studies 2, 393-421.
- Hansen, L., 1972, Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054.
- Haughton, K. and J. Christopherson, 1989, Equity Style Indexes: Tools for Better Performance Evaluation and Plan Management, Frank Russell Company White Paper, Tacoma, WA.
- Hedges, L.V. and I. Olkin, 1985, Statistical Methods for Meta-Analysis (Academic Press, Orlando, FL).
- Henriksson, R.D., 1984, Market Timing and Mutual Fund Performance: An Empirical Investigation, Journal of Business 57, 73-96.
- , and R.C. Merton, 1981, On the Market Timing and Investment Performance II: Statistical Procedures for Evaluating Forecasting Skills, Journal of Business 54, 513-533.
- Hsieh, D., 1983, A Heteroscedasticity-Consistent Covariance Matrix Estimator for Time Series Regressions, Journal of Econometrics 22, 281-290.
- Hunter, J.E. and T.D. Coggin, 1992, The Sampling Error in the Average Correlation, Working Paper, Michigan State University, East Lansing, MI.
- , T. D. Coggin, and S. Rahman, 1992, The Correlation Between Sampling Errors in the Estimation of Security Selection and Market Timing Ability, Working Paper, Michigan State University, East Lansing, MI.

- , and F.L. Schmidt, 1990, Methods of Meta-Analysis (Sage Publications, Newbury Park, CA).
- Ippolito, R.A., 1986, Pensions, Economics, and Public Policy (Dow Jones-Irwin, Homewood, IL).
- , 1989, Efficiency with Costly Information: A Study of Mutual Fund Performance, 1965-84, Quarterly Journal of Economics 104, 1-23.
- , and J.A. Turner, 1987, Turnover, Fees and Pension Plan Performance, Financial Analysts Journal 43, 16-26.
- Jagannathan, R. and R.A. Korajczyk, 1986, Assessing the Market Timing Performance of Managed Portfolios, Journal of Business 59, 217-235.
- Jensen, M.C., 1968, The Performance of Mutual Funds in the Period 1945-1964, Journal of Finance 23, 389-416.
- , 1969, Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios, Journal of Business 42, 167-247.
- , 1972, Optimal Utilization of Market Forecasts and the Evaluation of Investment Performance, in G.P. Szego and K. Shell, eds.: Mathematical Methods in Investment and Finance (North Holland, Amsterdam).
- Kon, S.J., 1983, The Market-timing Performance of Mutual Fund Managers, Journal of Business 56, 323-347.
- , and F.C. Jen, 1978, Estimation of Time-Varying Systematic Risk and Performance for Mutual Fund Portfolios: An Application of Switching Regression, Journal of Finance 33, 457-475.

- , and F.C. Jen, 1979, The Investment Performance of Mutual Funds: An Empirical Investigation of Timing, Selectivity, and Market Efficiency, Journal of Business 52, 263-289.
- Lakonishok, J., A. Shleifer, and R. W. Vishny, 1992, The Structure and Performance of the Money Management Industry, forthcoming in Brookings Papers on Economic Activity: Macroeconomics, Brookings Institution, Washington, DC.
- Lee, C.F. and S. Rahman, 1990, Market Timing, Selectivity, and Mutual Fund Performance: An Empirical Investigation, Journal of Business 63, 261-278.
- Lehmann, B. and D. Modest, 1987, Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons, Journal of Finance 42, 233-265.
- , and D. Modest, 1988, Empirical Issues in the Measurement of Market Timing Ability, Working Paper, University of California, Berkeley, CA.
- Tierney, D. E. and K. Winston, 1991, Using Generic Benchmarks to Present Manager Styles, Journal of Portfolio Management 17, 33-36.
- Trotman, K.T. and R. Wood, 1991, A Meta-Analysis of Studies on Internal Control Judgements, Journal of Accounting Research 29, 180-192.
- Thorndike, R.L., 1982, Applied Psychometrics (Houghton Mifflin, Boston, MA).
- Treynor, J.L. and K.K. Mazuy, 1966, Can Mutual Funds Outguess the Market?, Harvard Business Review 44, 131-136.
- White, H., 1980, A Heteroscedasticity-Consistent Covariance Matrix Estimator and A Direct Test for Heteroscedasticity, Econometrica 48, 817-838.

Table I

## Summary Statistics for Pension Manager Performance (1983-90)

This table presents summary statistics for the 71 managers for the entire sample period. The number of positive and negative selectivity and timing values are given for each model for each benchmark portfolio. Only those values which are significant at the .05 level or less are counted. The numbers in parentheses indicate the percent of the total sample.

|  | <u>Russell 3000</u> |                 | <u>Style Index</u> |                 | <u>S&amp;P 500</u> |                 |
|--|---------------------|-----------------|--------------------|-----------------|--------------------|-----------------|
|  | <u>Positive</u>     | <u>Negative</u> | <u>Positive</u>    | <u>Negative</u> | <u>Positive</u>    | <u>Negative</u> |
| <u>Bhattacharya and Pfleiderer Model</u> |                     |                 |                    |                 |                    |                 |
| Selectivity                              | 10(14.1%)           | 1(1.4%)         | 13(18.3%)          | 0(0%)           | 6(8.5%)            | 1(1.4%)         |
| Timing                                   | 2(2.8%)             | 4(5.6%)         | 3(4.2%)            | 5(7.0%)         | 1(1.4%)            | 6(8.5%)         |
| <u>Treynor and Mazuy Model</u>           |                     |                 |                    |                 |                    |                 |
| Selectivity                              | 9(12.7%)            | 2(2.8%)         | 13(18.3%)          | 0(0%)           | 6(8.5%)            | 1(1.4%)         |
| Timing                                   | 12(16.9%)           | 19(26.8%)       | 12(16.9%)          | 14(19.7%)       | 5(7.1%)            | 21(29.6%)       |



Table II

## Mean Values of Performance Measures Across Models and Benchmarks

This table presents the mean values of the selectivity and timing values across models and benchmarks for the entire sample period. The numbers in parentheses indicate the rank among investment styles for each measure.

|  | <u>Russell 3000</u> |               | <u>Style Index</u> |               | <u>S&amp;P 500</u> |               |
|--|---------------------|---------------|--------------------|---------------|--------------------|---------------|
|  | <u>Selectivity</u>  | <u>Timing</u> | <u>Selectivity</u> | <u>Timing</u> | <u>Selectivity</u> | <u>Timing</u> |
| <u>Bhattacharya and Pfleiderer Model</u> |                     |               |                    |               |                    |               |
| All Managers                             | .0008**             | -.0092        | .0016**            | -.0100        | .0003              | -.0470**      |
| Earnings Growth                          | -.0003 (3)          | .0538**(1)    | .0007* (4)         | .0217 (2)     | -.0008* (4)        | .0072 (1)     |
| Market Oriented                          | .0022**(1)          | .0012 (2)     | .0020**(2)         | -.0274 (3)    | .0017**(1)         | -.0451 (2)    |
| Price Driven                             | .0015**(2)          | -.0234 (3)    | .0009* (3)         | -.0843**(4)   | .0009* (2)         | -.0539**(3)   |
| Small Capitalization                     | -.0005 (4)          | -.0765**(4)   | .0031**(1)         | .0586* (1)    | -.0007 (3)         | -.1024**(4)   |
| <u>Treynor and Mazuy Model</u>           |                     |               |                    |               |                    |               |
| All Managers                             | .0008**             | -.0828        | .0016**            | -.0706        | .0004              | -.2799**      |
| Earnings Growth                          | -.0003 (3)          | .2014**(1)    | .0008* (4)         | -.0011 (2)    | -.0008* (4)        | .0108 (1)     |
| Market Oriented                          | .0021**(1)          | -.0278 (2)    | .0019**(2)         | -.1261 (3)    | .0017**(1)         | -.1799 (2)    |
| Price Driven                             | .0016**(2)          | -.0877 (3)    | .0009* (3)         | -.3286**(4)   | .0011**(2)         | -.2293* (3)   |
| Small Capitalization                     | -.0004 (4)          | -.4625**(4)   | .0032**(1)         | .2074* (1)    | -.0004 (3)         | -.7828**(4)   |

\*\* Significant at the .05 level

\* Significant at the .10 level

Table III

**Correlation of a Performance Measure Between Benchmarks\***

Each model was estimated for all managers for the entire period using each of the three benchmark portfolios. Panel A presents the Pearson and Spearman correlations between selectivity values for each pair of benchmark portfolios for each model. Panel B presents the Pearson and Spearman correlations between timing values for each pair of benchmark portfolios for each model.

## Panel A: Selectivity

|  | <u>Style Index</u> |                 | <u>S&amp;P 500</u> |                 |
|--|--------------------|-----------------|--------------------|-----------------|
|  | <u>Pearson</u>     | <u>Spearman</u> | <u>Pearson</u>     | <u>Spearman</u> |
| <u>Bhattacharya &amp; Pfleiderer Model</u> |                    |                 |                    |                 |
| Russell 3000                               | .806               | .744            | .997               | .994            |
| Style Index                                |                    |                 | .832               | .761            |
| <u>Treynor and Mazuy Model</u>             |                    |                 |                    |                 |
| Russell 3000                               | .804               | .734            | .996               | .995            |
| Style Index                                |                    |                 | .839               | .764            |

## Panel B: Timing

|  | <u>Style Index</u> |                 | <u>S&amp;P 500</u> |                 |
|--|--------------------|-----------------|--------------------|-----------------|
|  | <u>Pearson</u>     | <u>Spearman</u> | <u>Pearson</u>     | <u>Spearman</u> |
| <u>Bhattacharya &amp; Pfleiderer Model</u> |                    |                 |                    |                 |
| Russell 3000                               | .765               | .727            | .982               | .977            |
| Style Index                                |                    |                 | .777               | .739            |
| <u>Treynor and Mazuy Model</u>             |                    |                 |                    |                 |
| Russell 3000                               | .775               | .704            | .992               | .984            |
| Style Index                                |                    |                 | .721               | .627            |

\*All correlations are significant at the .0001 level

Table IV

**Correlation of a Performance Measure Between Models\***

Each model was estimated for all managers for the entire period using each of the three benchmark portfolios. This table presents the Pearson and Spearman correlations between selectivity values for each model for each benchmark, and the Pearson and Spearman correlations between timing values for each model for each benchmark.

| <u>Benchmark</u> | <u>Selectivity</u> |                 | <u>Timing</u>  |                 |
|------------------|--------------------|-----------------|----------------|-----------------|
|                  | <u>Pearson</u>     | <u>Spearman</u> | <u>Pearson</u> | <u>Spearman</u> |
| Russell 3000     | .992               | .988            | .901           | .923            |
| Style Index      | .991               | .990            | .835           | .930            |
| S&P 500          | .990               | .985            | .866           | .894            |

\*All correlations are significant at the .0001 level

Table V

**Correlation Between Selectivity and Timing**

Each model was estimated for all managers for the entire period using each of the three benchmark portfolios. This table presents the Pearson and Spearman correlations between the selectivity and timing values for each model for each benchmark.

| <u>Benchmark</u> | <u>Bhattacharya and Pfleiderer Model</u> |                    | <u>Treynor and Mazuy Model</u> |                    |
|------------------|--|--------------------|--------------------------------|--------------------|
|                  | <u>Pearson</u>                           | <u>Spearman</u>    | <u>Pearson</u>                 | <u>Spearman</u>    |
| Russell 3000     | -.447                                    | -.488              | -.485                          | -.427 <sup>a</sup> |
| Style Index      | -.359 <sup>d</sup>                       | -.315 <sup>c</sup> | -.399 <sup>b</sup>             | -.359 <sup>d</sup> |
| S&P 500          | -.487                                    | -.504              | -.467                          | -.387 <sup>c</sup> |

<sup>a</sup> significant at the .0002 level

<sup>b</sup> significant at the .0006 level

<sup>c</sup> significant at the .0008 level

<sup>d</sup> significant at the .0021 level

<sup>e</sup> significant at the .0075 level

All other correlations are significant at the .0001 level

Table VI

**Average Correlation Between Regression Residuals**

This table presents the average correlations between the regression residuals for the 71 managers for the entire period across benchmarks and models. The numbers in parentheses indicate the 95% confidence interval for the average correlation.\*

| <u>Benchmark</u> | <u>Bhattacharya and Pfeiderer Model</u> | <u>Treynor and Mazuy Model</u> |
|------------------|---|--------------------------------|
| Russell 3000     | .084 (.062, .123)                       | .083 (.061, .122)              |
| Style Index      | .071 (.052, .105)                       | .071 (.052, .105)              |
| S&P 500          | .232 (.177, .330)                       | .226 (.173, .321)              |

\* The formulas for the confidence interval of the average correlation are derived in Hunter and Coggin (1992). These formulas use the normal approximation to the chi-square distribution, and are reasonably accurate for sample sizes greater than 30. Let  $N$ =the number of time series returns per manager,  $n$ =the number of managers, and  $r$ =the average correlation between regression residuals. In our case,  $N=96$  and  $n=71$ . If we define  $X1 = 1.96 * \sqrt{2/(N-1)}$  and  $X2 = (1 + (n-1)*r)/(n-1)$ , then:

95% lower bound for  $r = r - [(X1/(1+X1)) * X2]$ , 95% upper bound for  $r = r + [(X1/(1-X1)) * X2]$ .

Table VII

## Meta-Analysis Results

This table presents the meta-analysis results for the selectivity and timing values based on the three benchmark portfolios and using heteroscedasticity-consistent standard errors, for the entire period (n=71 managers).

## Panel A: Bhattacharya and Pfeleiderer Model

|                  | Selectivity        |                     |                    | Timing             |                     |                    |
|------------------|--------------------|---------------------|--------------------|--------------------|---------------------|--------------------|
|                  | <u>S&amp;P 500</u> | <u>Russell 3000</u> | <u>Style Index</u> | <u>S&amp;P 500</u> | <u>Russell 3000</u> | <u>Style Index</u> |
| $\bar{b}$        | .000339            | .000769             | .001624            | -.046979           | -.009194            | -.010003           |
| $s_b$            | .002646            | .002646             | .002450            | .105010            | .114070             | .123984            |
| $s_{\beta}$      | .001827            | .001868             | .001807            | .054236            | .058340             | .075100            |
| $s_b^2$          | .000007            | .000007             | .000006            | .011027            | .013012             | .015372            |
| $s_{\beta}^2$    | .000003            | .000003             | .000003            | .002941            | .003404             | .005640            |
| $s_e^2$          | .000004            | .000004             | .000003            | .008086            | .009608             | .009732            |
| $\chi^2$ (df=70) | 130.66**           | 136.27**            | 145.78**           | 96.83**            | 96.15**             | 112.15**           |
| $s_e^2/s_b^2$    | .5714              | .5714               | .5000              | .7333              | .7384               | .6331              |

## Panel B: Treynor and Mazuy Model

|                  | Selectivity        |                     |                    | Timing             |                     |                    |
|------------------|--------------------|---------------------|--------------------|--------------------|---------------------|--------------------|
|                  | <u>S&amp;P 500</u> | <u>Russell 3000</u> | <u>Style Index</u> | <u>S&amp;P 500</u> | <u>Russell 3000</u> | <u>Style Index</u> |
| $\bar{b}$        | .000422            | .000796             | .001645            | -.279925           | -.082756            | -.070593           |
| $s_b$            | .002646            | .002646             | .002450            | .635032            | .598273             | .593544            |
| $s_{\beta}$      | .001922            | .001883             | .001927            | .588197            | .550055             | .552251            |
| $s_b^2$          | .000007            | .000007             | .000006            | .403266            | .357931             | .352295            |
| $s_{\beta}^2$    | .000003            | .000003             | .000003            | .345976            | .302560             | .304981            |
| $s_e^2$          | .000004            | .000004             | .000003            | .057290            | .055371             | .047314            |
| $\chi^2$ (df=70) | 145.24**           | 140.50**            | 165.14**           | 499.77**           | 458.96**            | 528.66**           |
| $s_e^2/s_b^2$    | .5714              | .5714               | .5000              | .1421              | .1547               | .1343              |

\*\* Significant at the .05 level or less.

Table VIII

**80% Probability Intervals for Observed and Population Selectivity and Market Timing Values**

This table presents the 80% probability intervals for the observed and population values of selectivity and market timing using all managers for the entire period. The observed values are bounded by  $\bar{b} \pm 1.28(s_b)$ , and the population values are bounded by  $b \pm 1.28(s_{\beta})$ .

**Panel A: Bhattacharya and Pfleiderer Model**

| Benchmark    | Observed Values |         |               |         | Population Values |         |               |         |
|--------------|-----------------|---------|---------------|---------|-------------------|---------|---------------|---------|
|              | Selectivity     |         | Market Timing |         | Selectivity       |         | Market Timing |         |
|              | Lower           | Upper   | Lower         | Upper   | Lower             | Upper   | Lower         | Upper   |
| S&P 500      | -.003121        | .003800 | -.181391      | .087434 | -.001999          | .002678 | -.116401      | .022443 |
| Russell 3000 | -.002687        | .004255 | -.155203      | .136814 | -.001623          | .003161 | -.083870      | .065481 |
| Style Index  | -.001606        | .004854 | -.168703      | .148697 | -.000689          | .003937 | -.106131      | .086126 |

**Panel B: Treynor and Mazuy Model**

| Benchmark    | Observed Values |         |               |         | Population Values |         |               |         |
|--------------|-----------------|---------|---------------|---------|-------------------|---------|---------------|---------|
|              | Selectivity     |         | Market Timing |         | Selectivity       |         | Market Timing |         |
|              | Lower           | Upper   | Lower         | Upper   | Lower             | Upper   | Lower         | Upper   |
| S&P 500      | -.003019        | .003864 | -1.092766     | .532916 | -.002038          | .002883 | -1.032817     | .472967 |
| Russell 3000 | -.002632        | .004223 | -.848546      | .683034 | -.001615          | .003206 | -.786826      | .621314 |
| Style Index  | -.001622        | .004912 | -.830330      | .689144 | -.000822          | .004112 | -.777474      | .636288 |











Date Due

Lib-26-67

MIT LIBRARIES



3 9080 00779671 4

